DOCUMENT RESUME

ED 160 095 IR 006 342

AUTHOR Heller, Jack
TITLE On Logical Data Organization, Card Catalogs, and the GRIPHOS Management Information System. Museum Data Bank Research Report No. 3.
INSTITUTION California State Univ., Fullerton. Div. of Library Science.
PUB DATE Dec 74
NOTE 23p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Catalogs; Data; Electronic Data Processing; Humanities; *Information Retrieval; *Management Information Systems; Mathematical Models; Objectives; *Organization
IDENTIFIERS *GRIPHOS

ABSTRACT

This report presents the rationale, motivation, and formal description underlying the General Retrieval and Information Processor for Humanities Oriented Studies (GRIPHOS) Management Information System. This system was devised to handle "card catalog" types of information; perform searches that are usually carried out in archives, libraries and museums; allow for the maintenance and updating of records; and produce reports such as frequency dictionaries, union catalog listings, statistical analysis, and selective print-outs. An example of a complex card catalog which contains hierarchical structures and one step threading is presented to illustrate the types of data structures that have been used historically to construct and maintain catalogs. A formal description of the implementation of GRIPHOS is presented in set and graph theoretic terms, and the logical structure of data as conceived by users is described in standard mathematical notation. (Author/MBR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED160095

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MUSEUM DATA BANK RESEARCH REPORT

Number 3

On Logical Data Organization, Card Catalogs, and the GRIPHOS Management Information System

by

Jack Heller

December 1974

This report, or any portion thereof, may be reproduced for any non-commercial purpose provided appropriate credit is given to the author and the Museum Data Bank Committee as publisher.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. G. Chenhall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Museum Data Bank Committee
Robert G. Chenhall, Chairman

Margaret Woodbury Strong Museum
700 Allen Creek Road
Rochester, N.Y.

1 R006 342

Introduction. This report presents the rationale, motivation and formal description underlying the GRIPHOS management information system.[1] The GRIPHOS system was devised to handle "card catalog" type of information, perform the searches that are usually carried out in archives, libraries and museums, allow for the maintenance and updating of records and produce reports such as frequency dictionaries, union catalog listings, statistical analysis and selective print outs.

In section 2 we give an example of a complex card catalog which contains hierarchal structures and one step threading. The purpose of this example is to illustrate the types of data structures that have historically been used by those who construct and maintain card catalogs.

In sections 3 and 4 we present a formal description of the present implementation of GRIPHOS (Heller n.d.; Lowy 1974; Chenhall 1975) in set and graph theoretic terms. Our view of a description is to use - wherever possible - a standard mathematical notation to describe the type and structure of data in line with the philosophical views of Codd's (1973) relational data set model. We view the logical structure of data as conceived by the user and describe this concept by standard mathematical notation so that we can use the power of the mathematics to analyse the user's understanding and organization of the data. The question of an implementation is treated elsewhere (Heller n.d.; Lowy 1974). We show that one of the query, selection, relation sequence of procedures in GRIPHOS results in the construction of relations as viewed by Codd (1973), while another query, selection, relation produces a relation that is ruled out by Codd's assumption of column homogeneity in the relational model.

2 Intuitive Example. In this section we will give an example of a data organization that mimics to some extent card catalogs and computerized data bases treated by the GRIPHOS management information system. For this example we have chosen "records" of a few art objects, art reference books, artist's biographies and museums. Admittedly this example is simplistic but contains the logical aspects of data organization which is formally described in sections 3 and 4 using a set theoretic notation.

The data that we will consider is textual in nature and is seemingly unstructured. Bibliographic information collected and accumulated in libraries, museums, archives, government agencies, academic institutions, employment divisions of commercial establishments, ... is typical of data that we are considering. This type of data is unordered and variable in amount, size and structure. Because of these properties and the state of the art in using data processing devices, it is necessary to tag each piece of information or require some fixed format and typographical display. We choose a tagging concept.

---

1. The acronym GRIPHOS was first suggested by Professor Alice Pollin and stands for General Retrieval and Information Processor for Humanities Oriented Studies.

A record is viewed as a card catalog entry containing varying numbers of fields[2] of information composed of tag, value and connectivity data[3]. We will visualize these records as lists of fields in which the tag is in the first column, the value is in the second column, and the connectivity information, if it exists, is in the third column. Each record will be given a name of the form RECORD number and will appear at the head of the list of fields for that record. In order to visualize the records we will place a box around each record, and when necessary subdivide each box. The collection of records or data base for the present discussion is displayed in Appendix I.

RECORD 2 is a typical simple record containing fields with tags TYPE, ACCNO, ARTIST, TITLE, MADE, SIZE, LOC, REF, and MEDIUM. Similarly, the other art objects referred to in the data base are in records 1-7 and 22 containing tags, values and connectivity information.

The connectivity information designated by T- (.,.) is a convenience of data storage and maintenance. Each art object recorded in this data base contains fields (ARTIST, REF, LOC) which have related data common to many art objects, e g. an artist's biography related to ARTIST and name, a bibliographic entry related to REFerence and the Library of Congress Number or local call number; a museum's location and structure related to LOCation of art object and museum's name. This connectivity is designated in a field by the syntax *T- (.,.) where (.,.) gives the number of the record and if need be its subpart. For example, in RECORD 2, the field whose tag is REF contains the Library of Congress Number value 53-11065 and is related to RECORD 9 via *T- (9,). In RECORD 9, we see that there is an AUTHOR, TITLE, PUBlication, Publication DATE, TRANSlator, ... This document, i.e. REFNO 53-11065 is also referred to by RECORD 1 and RECORD 4. The main advantage of this cross referencing of records is that there is less storage and maintenance required as the data base grows. For example, if an addition to the bibliographic information of a reference is required, it need only be placed in the reference record: all connected information will remain unaltered.

This type of organization is not uncommon in card catalogs in which a field in a record may refer to another record implicitly by its value, requiring the user, if desired, to look up this new record. For example, a record of an art object may contain a designation SHOW with its value name and number. This name and number can be used to look up the record of the show, usually the brochure, in which there is detailed information about the show. Many art object records in the card catalog can point to the same show. This concept was used by Vance (Vance and Heller 1971) and Martini (Martini 1972) in the use of the GRIPHOS management information system. In the forthcoming book by Chenhall (Chenhall 1975) this concept is explicitly

---

2. A field of information as we use it is referred to by many different names: unit, atom, bibliographic item, ...

3. There are many different forms of jargon for tag: name as used in CODASYL reports, identifier as used by the United Nations Indexing Project, annotation class as used by the Museum Computer Network, domain as used in the relational data set model, ...

stated as a desirable strategy for data collection.

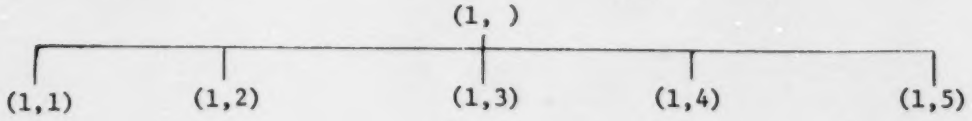
Queries about the information on these "cards" can be answered by searching through each record and determining if the information on the card satisfies the query. Typical manual card catalogs found in libraries and museums are ordered so as to facilitate the answering of queries. If queries are directed to finding a given artist, the cards are ordered lexically by artist; if the queries are directed to finding art objects in a given medium, the cards are ordered by medium Sometimes cards are duplicated and stored intermixed under various terms which may arise in a class of queries. In some computer based card catalogs the records are stored in the order that they are prepared and entered into the data base. Therefore in a computer based card catalog, it is necessary to generate various aids with the management information system to facilitate the answering of queries. In Appendix II we have illustrated an inverted file type of index to part of the data base of Appendix I.

Let us consider a query and search the cards sequentially. Consider the query: what oil paintings are found in the data base? Searching for a match on TYPE oil painting, we find that RECORDS 1, 2, 4, 5, and 7 satisfy the query. We are also convinced, at this point, that by matching tag and value we will be able to answer a query.

Let us consider a more complex query: what oil paintings contain a musician? Searching for a match on (TYPE oil painting) and (FIGURE musician), we find that only RECORD 5 contains a match on these two fields. Again we are convinced that matching a tag and value within a record will be able to answer a query.

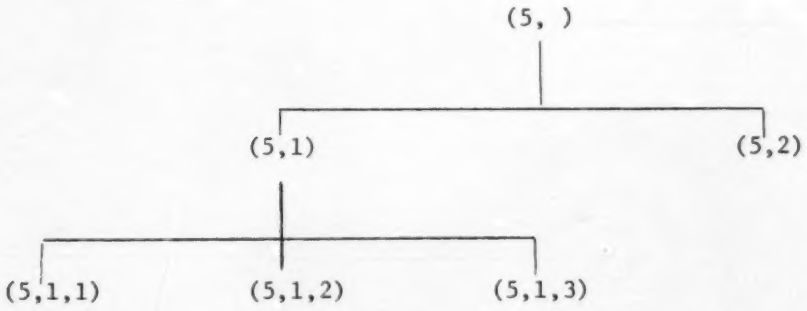
Let us consider another complex query: what oil paintings contain a musician with a clarinet and music score? Searching within each record for a match on (FIGURE musician) and (OBJECT clarinet) and (OBJECT music score), we find RECORD 5 containing these 3 fields. However, the answer to the query is not correct: we wanted a musician with a clarinet and music score, not a musician with a clarinet and another musician with a music score.

In order to resolve these false drops we find it necessary to organize a record into an hierarchal (tree) structure where all the fields along a path from a bottom (leaf) node to the apex (root) node are logically connected for the set of queries that we are considering. For example in RECORD 1 we have the hierarchal structure



We have indicated this structure by subdividing the record into parts (partitions), labeling these parts (1,1), (1,2), (1,3) (1.4) and (1,5). Also we have indicated the "up" connection of a part of a record by *U- (1,1), *U- (1,2), etc., and have let the apex part be designated by (1,). We note that the FIGURE described in (1,1) has position left and the FIGURE described in (1,2) has a position front and seated. There is no confusion about which figure is facing front or left or seated or painting or seated with head in arms at table, ...

RECORD 5 has a more complex hierarchal structure:



Since there are three FIGURES in the painting which are musicians, it is possible to place the field FIGURE musician in a node from which emanates the nodes (5,1,i), (5,1,2) and (5,1,3) containing fields about <u>each</u> musician.

There is a relatively simple procedure by which an hierarchal structure can be formed:

(1) place general overall descriptive fields in the root partition,

(2) subdivide the remaining properties of the object into their logical subparts

and (3) for each subpart repeat (1) and (2) until all fields are placed in their respective partitions of the record.

As an example, let us consider a portfolio of three drawings that are to be considered as one record. The title of this portfolio is "Children" and each art object is titled respectively "Boys", "Girls" and "Young Girl". The first two objects were executed by Blyden, John and the remaining object was constructed by Heller, Glen. There are 6 partitions of the RECORD 22 to reflect the interrelations of the fields in this composite record of 3 art objects.

GROUP	Portfolio
ACCNO	p70.5A-C
TYPE	drawing
ARTIST	Blyden, John
SIZE	5 x 7"
ACCNO	p70.5A
TITLE	Boys
ACQUIRE	Gift of the artist 1970
ACCNO	p70.5B
TITLE	Girls
DONOR	Heller, Glen
ARTIST	Heller, Glen
ACCNO	p70.5C
TITLE	Young Girl
SIZE	8 1/2 x 11"
VENDOR	Heller, Glen
TITLE	Children
VENDOR	Jones, R. S., and Co.

Following out step (1) in the above algorithm, we take as the root fields

GROUP	Portfolio
ACCNO	p70.5A-C
TYPE	drawing

Step (2) of the algorithm suggests the subdivisions of the non-root information:

ARTIST	Blyden, John
SIZE	5 x 7"
ACCNO	p70.5A
TITLE	Boys
ACQUIRE	Gift of the artist 1970
ACCNO	p70.5B
TITLE	Girls
DONOR	Heller, Glen

ARTIST	Heller, Glen
ACCNO	p70.5C
TITLE	Young Girl
SIZE	8 1/2 x 11"
VENDOR	Heller, Glen

TITLE	Children
VENDOR	Jones, R. S., and Co.

Repeating steps (1) and (2) for the first subdivision gives us the partitions:

ARTIST	Blyden, John
SIZE	5 x 7"

ACCNO	p70.5A
TITLE	Boys
ACQUIRE	Gift of the artist 1970

ACCNO	p70.5B
TITLE	Girls
DONOR	Heller, Glen

Continuing on in this manner, we arrive at the subdivided RECORD 22 given in APPENDIX I. We can view this structure in the following hierarchal form:

(22,)

GROUP	portfolio
ACCNO	p70.5A-C
TYPE	drawing

(22,1)

ARTIST	Blyden, John
SIZE	5 x 7"

(22,2)

ARTIST	Heller, Glen
ACCNO	p70.5C
TITLE	Young Girl
SIZE	8½ x 11"

(22,3)

TITLE	Children
VENDOR	Jones, R.S. & Co.

(22,1,1)

ACCNO	p70.5A
TITLE	Boys
ACQUIRE	Gift of the artist 1970

(22,1,2)

ACCNO	p70.5B
TITLE	Girls
DONOR	Heller, Glen

If we were to ask the query, what art objects were donated by Heller, Glen and we searched all the records from their leaf nodes upward, we would answer the query correctly by searching those nodes only for the desired set of fields. The answer to the query would give the fields:

ACCNO	p70.5B
TITLE	Girls
DONOR	Heller, Glen
ARTIST	Blyden, John
SIZE	5 x 7"
GROUP	portfolio
ACCNO	p70.5A-C
TYPE	drawing

3 Formal Description of GRIPHOS. The intuitive example of a logical data base organization given in section 2 can be clarified by stating the nature of the data in a formal mathematical description. We find it convenient to use a set and graph theoretic description of logical data that we deal with in GRIPHOS, with some undefined terms which are given a realization by a specific implementation (e.g., as described in Lowy 1974). The undefined terms are tag, value, partition and upward pointer and all other terms are defined as sets and operations on these sets. In section 2 we designated tags as ARTIST, ACCNO, SIZE, NAME, ...; values as Miro, Jean, 49.21S, Modern Painting, 5 x 7", Picasso, Pablo, ...; partitions as (1,1), (1,2), (5,1,3), ...; upward pointers as *U- (5,1), *U- (5,1,1), *U- (1,), ...

In our formal description we define a tag as an element of the set[4] .

$$T=\{t_\sigma / \sigma=1,\ldots,\#T\} \tag{3.1}$$

a value as an element of the set

$$V=\{v_\sigma / \sigma=1,\ldots,\#V\} \tag{3.2}$$

a partition name as an element of the set

$$H=\{h_\sigma / \sigma=1,\ldots,\#H\} \tag{3.3}$$

and an upward pointer as an element of the set

$$U=\{u_\sigma / \sigma=1,\ldots,\#U\} \tag{3.4}$$

In terms of these undefined quantities we define a field as an ordered 4-tuple.

$$f=\langle t,v,h,u\rangle \tag{3.5}$$

A record is defined as a set of fields

$$r=\{f_\sigma / \sigma=1,\ldots,\#r\} \tag{3.6}$$

and a data base is defined as a set of records

$$D=\{r_\sigma / \sigma=1,\ldots,\#D\} \tag{3.7}$$

The user specifies the fields which make up a record. The choice of t and v are by historical convention and usage; the choice of h and u are made by the user to resolve "false drops" with respect to the queries that he desires to carry out.

In order to define the query set supported by GRIPHOS we find it

---

4. We will use Greek subscripts to indicate an index of a set, lower case Latin subscripts to indicate a specific indexed member of a set and lower case Latin characters to indicate a member of a set.

convenient to define certain operators. We define Λ_i which operates on an ordered n-tuple giving us the ith element of the n-tuple. For example, the f defined in (3.5),

$$\Lambda_1(f) = t$$

$$\Lambda_2(f) = v$$

$$\Lambda_3(f) = h$$

$$\Lambda_4(f) = u$$

The h s in a set of fields of a record r_i form a partition of r_i which we write as

$$P_{ij} = \{f \mid f \in r_i \& \Lambda_3(f) = h_j\} \tag{3.8}$$

If we define Γ_u as the mapping between partitions, the partitions of a record in GRIPHOS satisfy

(1) $\Gamma_u P_u = \phi$,

(2) for every P_{ij}, $\Gamma_u P_{ij} = P_{ij'}$ $j = 1$,

and (3) for every P_{ij}, $\exists\, n \ni \Gamma_u^n P_{ij} = \phi$.

The graph of the partitions of a record r_i

$$R_i = [\Gamma_u, \{P_{io} \mid p_i\Gamma \in r_i\}] \tag{3.9}$$

is an arborescence (Berge 1962). Condition (1) implies that there is a root to the graph; condition (2) implies that there is no circuit to the graph; condition (3) implies that the graph is connected. The GRIPHOS system uses as the data structure of a record an arborescence and as a data base a network of arborescences. It is believed that only this structure is necessary for the general queries that are asked of data bases; it is believed that the general networks allowed in CODASYL (1971) are not necessary.

One of the purposes of a data management system is to organize data and disseminate subsets of this data on request. The organization of the data depends on the class of queries and the user's concept of a correct relation. We will pursue the concept of a query and the concept of a correct relation defining these concepts in terms of the above defined sets of data.

First we define a leaf node as a partition P_{ij} such that $\Gamma_u^1 P_{ij} = \phi$. For each leaf node, since each record is an arborescence of partitions there is a unique sequence of partitions P_{ij}, P_{ij_1}, P_{ij_2}, ..., P_{i_1} given by the mappings $\Gamma_u^0 P_{ij}$, $\Gamma_u^1 P_{ij}$, ..., $\Gamma_u^n P_{ij}$. We will define Γ_u^* by $\Gamma_u^* P_{ij} = \{P_{ij}, \Gamma_u P_{ij}, \Gamma_u^2 P_{ij}, \ldots\}$ (3.10)
Using this notation, we define the leaf paths of a record as

$$l_{ij} = \{\Gamma_u^* P_{ij} \mid \Gamma_u^1 P_{ij} = \phi\} \tag{3.11}$$

The subsets l_{ij} of a record form a covering of the record: $\cup_j P_{ij} = r_i$. The user in forming the "logically relatable" subsets of fields in a record gives the partitions P_{ij} and their connectivity. By this partitioning and connectivity, he is stating those fields that can be "added" in a query and give correct relations. The relational data set concept of Codd

(1973) assumes that these relations are given. It is by no means easy to give all possible relations in a complex data base[5]; it is the data management system operating on a data base that should produce the correct relations at query time. Because of the possibility of "false drops" and "connection traps" we must specify, in detail, the query classes and the meaning of a correct relation.

The relating of one set of data to another is supported in the GRIPHOS system by a "one step thread". This "one step thread" is defined by the mapping $\Gamma_{tt'}$ of a field to a partition

$$\Gamma_{tt'}\, f = \{p_{\omega\sigma} \mid \exists f' \in p_{\omega\sigma}\ \Pi_1(f') = t' \notin \Pi_2(f) = \Pi_2(f') \quad (3.12)$$
$$\notin \Pi_1(f) = t \notin \langle \Pi_1(f'), \Pi_2(f') \rangle \text{ is unique} \}$$

There is a third mapping Γ_r supported by the GRIPHOS system which operates on the root partition of a record;

$$\Gamma_r\, p_{i1} = p_{i+1,1} \qquad i < \#D \qquad (3.13)$$
$$= \phi \qquad i = \#D$$

From the point of view of the mapping Γ_u a record is an arborescence; from the point of view of the mapping Γ_r a GRIPHOS data base is a one way list of arborescences; from the point of view of the mapping $\Gamma_{tt'}$ a GRIPHOS data base is a network of arborescences. If we view the data bank as a set of nodes (partitions), there are labels to the branches: u, tt' and r for the mappings. Γ_u, $\Gamma_{tt'}$ and Γ_r. (Schneiderman 1972)

There is another structure supported by GRIPHOS: an "inverted file" structure. In order to define this structure, we find it convenient to define (CODASYL 1973)

$$g = \langle \Pi_1(f), \Pi_2(f) \rangle = \Pi_{1,2}(f) = \langle t, v \rangle \qquad (3.14)$$

Considering all the g s in a data base, we define the set

$$G = \{g_\sigma \mid \sigma = 1, \ldots, \#G\} \qquad (3.15)$$

Partitioning the set G by t, we define

$$G_t = \{g_\sigma \mid g_\sigma \in G \notin \Pi_1(g) = t_t\} \qquad (3.16)$$

We define a "hash function" operating on v :

$$10_{b_t}(v) = m \qquad (3.17)$$

where m is an integer in the closed interval $[1, b_t]$ and the b are the so called number of buckets that the hash function uses to store those

---

5. For example in the Museum Computer Network's data base there are upwards of 1500 tags (Vance and Heller 1971)

v s which appear in the data base with tag t_γ. With this hash function, we partition G :

$$G_{\gamma m} = \{ g_\sigma \mid g_\sigma \in G_\gamma \ \& \ 10_b \Pi_1(g)) = m \} . \tag{3.18}$$

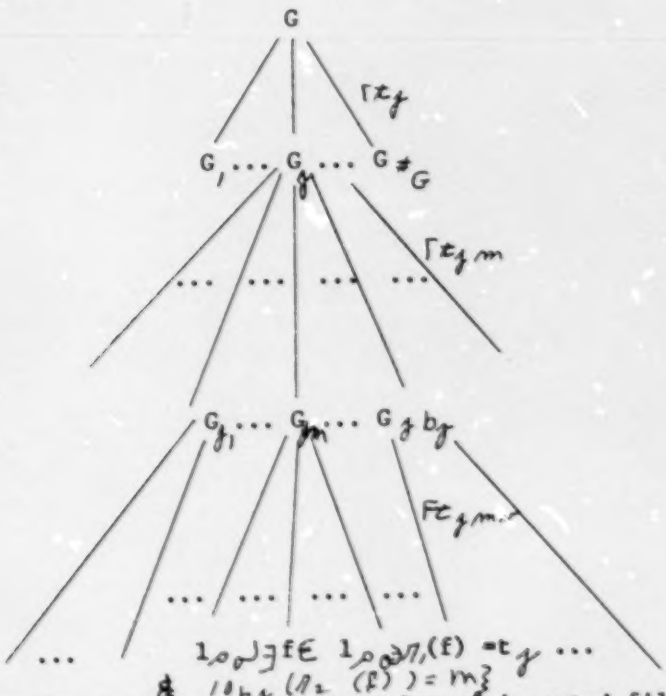
The structure of the inverted file supported by GRIPHOS searches the sets G, G_γ and then $G_{\gamma m}$ to give the leaf paths containing and f $\exists \langle \Pi_1(f), \Pi_2(f) \rangle \in G_{\gamma m}$. The mappings corresponding to this search are $\lceil t_\gamma$, $\lceil t_{\gamma m}$ and $\lceil t_\gamma m \nu$

$$\lceil t_\gamma \ G = \{ G_\gamma \} \tag{3.19}$$

$$\lceil t_\gamma m G = \{ G_{\gamma m} \} \tag{3.20}$$

$$\lceil t_\gamma m \nu \ G_{\gamma m} = \{ 1\rho_\sigma \mid \exists f \in 1\rho_\sigma \ni(\Pi_1(f), \Pi_2(f) \rangle \in G_{\gamma m} \} \tag{3.21}$$

A structure of this type is called "tree searchable" as can be seen from the diagram:


G
$\lceil t_\gamma$
$G_1 \ldots G_\gamma \ldots G_{\neq G}$
$\lceil t_{\gamma m}$
$G_{\gamma 1} \ldots G_{\gamma m} \ldots G_{\gamma b_\gamma}$
$\lceil t_{\gamma m \nu}$
$1\rho_\sigma \exists f \in 1\rho_\sigma \Pi_1(f) = t_\gamma \ldots$
$\& \ 10_{b_\gamma}(\Pi_2(f)) = m\}$

The GRIPHOS system supports the construction of inverted files of this type for any subset of T and bucket sizes b_γ, one for each t in the subset of T.

4 Query, Selection, Relation. By searching the data structures formally described in section 3, we are able to answer queries, select subsets of fields and construct relations among the selected data. So far, the user has grouped his data in logically connectable sets of fields: it is necessary that the searches through these connected sets produce relations that fall within these user defined logically connectable sets. We will define the queries presently

supported by GRIPHOS, and the generation of relations from selections.

The user has defined leaf paths as logically connectable sets of fields (LCSF) and a field with its thread path as another LCSF. The query must reflect this grouping of fields. We define a simple query taking this grouping into account:

$$Q(g,y) = \{l_{\rho\sigma} \exists f \in l_{\rho\sigma} \exists \Pi_{12}(f) = g \not\in g \; y \; \Pi_{12}(f)\} \tag{4.1}$$

where the binary relation designated by y is one of the following:

(1) exact match (=")
(2) substring match (=)
(3) leading substring match (<>)
(4) lexically less than (<)
(5) lexically less than or exact match (<=)
(6) lexically greater than (>)
(7) lexically greater than or exact match (>=)
(8) not exact match (¬=).

In GRIPHOS compound queries involving "and" and "or" are intersections and unions of simple queries.

Once a query is performed, a selection based on an ordered set of tags and a leaf path will give rise to a set of fields from which relations are to be constructed. We define a selection as

$$\delta(l_{\rho\sigma}, \langle t_{k_1}(t_{m_1}), t_{k_2}(t_{m_2}), \ldots t_{k_m}(t_{m_m})\rangle) = \langle S_1, S_2, \ldots S_m\rangle \tag{4.2}$$

where

$$S_m = \{\langle f,f'\rangle \mid f \in l_{\rho\sigma} \not\in \Pi_1(f) = t_{k_m} \not\in f' \in (l_{\rho\sigma} \mid [t_{k_m} t_{\ell_m}] \not\in \Pi_1(f') = t'_{\ell_m}\} \tag{4.3}$$

There are two types of relations supported by GRIPHOS, one is the relation defined by Codd(1973) and the other ruled out by Codd because of the non allowability of more than one field with the same tag in a leaf path. A Codd type relation set is

$$R_C \mid \delta(l_{\rho\sigma}, \langle t_{k_1}(t_{\ell_1}), t_{k_2}(t_{\ell_2}), \ldots, t_{k_m}(t_{\ell_m})\rangle) = C_1(v_{k_1}) \times C_2(v_{k_2}) \times \ldots \times C_m(v_{k_m}) \tag{4.4}$$

where

$$C_m(v_{k_m}) = \{f,f' \mid \langle f,f'\rangle \in S_m \not\in \Pi_2(f) = v_{k_m}\} \tag{4.5}$$

The other relation presently supported by GRIPHOS is

$$R_C(\delta(l_{\rho\sigma}, \langle t_{k_1}(t_{\ell_1}), t_{k_2}(t_{\ell_2}), \ldots t_{k_m}(t_{\ell_m})\rangle) = \langle S_1, S_2, \ldots, S_m\rangle \tag{4.6}$$

APPENDIX I

RECORD 1

TYPE	oil painting	
ACCNO	35.12X	
TITLE	The Muse	
ARTIST	Picasso, Pablo	*T->(16,)
MADE	1935	
SIZE	51 1/4 x 63 3/4"	
LOC	Musee D'Art Moderne	*T->(15,)
REF	53-11065	*T->(9,)
MEDIUM	oil	
(1,1)	*U->(1,)	
FIGURE	woman	
POSITION	left	
COMMENT	seated on cushion, facing left and painting	
(1,2)	*U->(1,)	
FIGURE	woman	
POSITION	front	
COMMENT	seated head in arms at table, facing front	
POSITION	seated	
(1,3)	*U->(1,)	
OBJECT	table	
(1,4)	*U->(1,)	
OBJECT	vase	
COMMENT	vase on table with flowers	
(1,5)	*U->(1,)	
OBJECT	mirror	

RECORD 2

TYPE	oil painting	
ACCNO	12.263p	
ARTIST	Klee, Paul	*T->(19,)
TITLE	Villa R	
MADE	1919	
SIZE	10 1/8 x 8 3/4"	
LOC	Kunstmuseum, Basel	*T->(14,)
REF	53-11065	*T->(9,)
MEDIUM	oil	

RECORD 3

TYPE	Sculpture	
ACCNO	51.302s	
TITLE	Baboon and Young	
ARTIST	Picasso, Pablo	*T->(16,)
FUND	Mrs. Simon Guggenheim Fund	
MADE	1951	
LOC	The Museum of Modern Art, New York	*T->(13,)
SIZE	21" high	
REF	67-29295	*T->(8,)
REF	70-164877	*T->(11,)

RECORD 4

TYPE	oil painting	
ACCNO	private 36	
ARTIST	Braque, Georges	*T->(20,)
TITLE	Patience	
MADE	1942	
SIZE	57 1/4 x 44 1/2"	
LOC	Private collection, U.S.A.	
REF	53-11065	*T->(9,)
MEDIUM	oil	

RECORD 5

TYPE	oil painting	
ACCNO	21.37p	
TITLE	The Three Musicians	
ARTIST	Picasso, Pablo	*T->(16,)
VENDOR	Gallerie Rosenberg	
MADE	1921	
LOC	The Museum of Modern Art, New York	*T->(13,)
SIZE	79 x 87 3/4"	
REF	Our-2	*T->(12,)
REF	70-164877	*T->(11,)
MEDIUM	oil	
MEDIUM	canvas	

(5,1)	U- (5,)
FIGURE	musician
POSITION	front

(5,1,1)	U- (5,1,1)
OBJECT	clarinet
COMMENT	left figure

(5,1,2)	U- (5,1)
OBJECT	guitar
COMMENT	center figure

(5,1,3)	U- (5,1)
OBJECT	accordian
OBJECT	music score
COMMENT	right figure

(5,2)	U- (5,)
FIGURE	dog
POSITION	side

RECORD 6

TYPE	Sculpture	
ACCNO	49.21S	
ARTIST	Picasso, Pablo	* T->(16,)
TITLE	Owl	
MADE	1949	
LOC	Galerie Chalette, New York	
SIZE	7 7/8 x 8 1/4 x 5 1/4"	
REF	67-29395 57 68 T- (8,)	
MEDIUM	ceramic	

RECORD 7

TYPE	oil painting	
ACCNO	05.21	
ARTIST	Matisse, Henri	*T->(17,)
TITLE	Luxe, Calme et Volupte	
DONOR	J. Rump	
MADE	1905	
LOC	Royal Museum of Fine Art, Copenhagen	
SIZE	82 3/4 x 54 1/4"	
REF	Our-1	*T->(10,)
REF	53-11065	*T->(9,)
MEDIUM	oil	

RECORD 8

REFNO	67-29395	
PUBL	The Museum of Modern Art, New York	*T->(13,)
AUTHOR	Penrose, Roland	
TITLE	The Sculpture of Picasso	
PDATE	1967	
NOPGS	231	
NOILLUST	489	

RECORD 9

REFNO	53-11065
AUTHOR	Raynal, Maurice
TITLE	Modern Painting
PUBL	SKIRA
PDATE	
TRANS	Gilbert, Stuart
PRINTED	Switzerland
NOPGS	339
NOILLUST	258

RECORD 10

REFNO	Our-1
AUTHOR	Raynal, Maurice
AUTHOR	Rudlinger, Arnold
AUTHOR	Bollinger, Hans
AUTHOR	Lassaigne, Jacques
TRANS	Gilbert, Stuart
PRINTED	Switzerland
NOPGS	149
NOILLUST	98
PUBL	SKIRA
PDATE	1950

RECORD 11

REFNO	70-164877
AUTHOR	Rubin, William
AUTHOR	Johnson, Elaine L.
PUBL	The Museum of Modern Art, New York * T->(13,)
AUTHOR	Castleman, Riva
PDATE	1972
DISTRB	New York Graphic Society Ltd.
NOPGS	246
NOILLUST	318

RECORD 12

REFNO	Our-2
AUTHOR	Cassou, Jean
TITLE	Picasso
PUBL	Editions Hyperon, France
PDATE	1940
NOPGS	167
NOILLUST	146

RECORD 13

MUSEUM	The Museum of Modern Art, New York
FROM	1929
LOC	11 West 53 Street, New York, New York 10019, U.S.A.
TYPE	private

RECORD 14

MUSEUM	Kunstmuseum, Basel
LOC	Basel, Switzerland

RECORD 15

MUSEUM	Musee D'Art Moderne, Paris
LOC	Paris, France

RECORD 16

NAME	Picasso, Pablo
BORN	1881
DIED	1973
BIOG	Spanish, born 1881. In France since 1904. Died 1973
COUNTRY	Spain
LIVED	France
LIVED	Spain

RECORD 17

NAME	Matisse, Henri
COUNTRY	France
BORN	1869
DIED	1954
BIOG	France, 1869-1954
LIVED	France

RECORD 18

NAME	Miro, Joan
COUNTRY	Spain
BORN	1893
BIOG	Spanish, Born 1893. Lived in Paris 1919-1940.
LIVED	England
LIVED	Spain

RECORD 19

NAME	Klee, Paul
COUNTRY	Germany
BORN	1879
DIED	1940
BIOG	German, 1879-1940. Born and died in Switzerland.
LIVED	Germany
LIVED	Switzerland

RECORD 20

NAME	Braque, Georges
COUNTRY	France
BORN	1882
DIED	1963
LIVED	France
BIOG	French, 1882-1963

RECORD 21

NAME	Heller, Glen
COUNTRY	U.S.A.
BORN	1950
LIVED	U.S.A.
LIVED	England
BIOG	U.S.A., Born 1950, Lived in England 1969-72

RECORD 22

GROUP	Portfolio
ACCNO	p70.5A-C
TYPE	drawing

(22,1) U- (22,)

ARTIST	Blyden, John
SIZE	5 x 7"

(22,1,1) U- (22,1)

ACCNO	p70.5A
TITLE	Boys
ACQUIRE	Gift of the artist 1970

(22,1,2) U- (22,1)

ACCNO	p70.5B	
TITLE	Girls	
DONOR	Heller, Glen	*T->(21,)

(22,2) U- (22,)

ARTIST	Heller, Glen	*T->(21,)
ACCNO	p70.5C	

TITLE	Young Girl	
SIZE	8 1/2 x 11"	
VENDOR	Heller, Glen	*T->(21,)

(22,3) U- (22,)

TITLE	Children
VENDOR	Jones, R. S., and Co.

APPENDIX II

An inverted file of the data base of APPENDIX I on tags ARTIST, FIGURE, LOC, and TYPE.

ARTIST	Blyden, john	(22,1,1) (22,1,2)
	Braque, Georges	(4,)
	Heller, Glen S.	(22,2)
	Klee, Paul	(2,)
	Matisse, Henri	(7,)
	Picasso, Pablo	(1,1) (1,2) (1,3) (1,4) (1,5) (3,) (5,1,1) (5,1,2) (5,1,3) (5,2)
FIGURE	dog	(5,2)
	musician	(5,1,1) (5,1,2) (5,1,3)
	women	(1,1) (1,2)
LOC	Gallerie Chalette, New York	(6,)
	Kunstmuseum, Basel	(2,)
	Musee D'Art Moderne, Paris	(1,1) (1,2) (1,3) (1,4) (1,5)
	The Museum of Modern Art, New York	(1,) (5,1,1) (5,1,2) (5,1,3) (5,2)
	Private Collection, U.S.A.	(4,)

TYPE	drawing	(22,1,1) (22,1,2) (22,2) (22,3)
	oil painting	(1,1) (1,2) (1,3) (1,4) (1,5) (5,1,1) (5,1,2) (5,1,3) (5,2) (7,)
	sculpture	(3,)

REFERENCES CITED

Berge, C.

1962 The Theory of Graphs and Their Applications. John Wiley and Sons, Inc., New York.

Chenhall, R. G.

1975 Museum Cataloging in the Computer Age. American Association for State and Local History, Nashville, Tennessee.

CODASYL

1973 Data Descriptive Language, Journal of Development, NBS Handbook 113.

Codd, E. F.

1973 A Relational Model of Data for Large Shared Banks. CACM, V. 13, no. 6.

Data Base Task Group to the CODASYL Programming Committee

1971 ACM Report

Heller, J.

n.d. GRIPHOS: General Retrieval and Information Processor for Humanities Oriented Studies.

Lowy, H.

1974 Architecture of Information Systems.

Martini, G. S.

1972 The Computer-Assisted Indexing Programme of the United Nations. Internal Report.

Schneiderman, Ben

1972 A Graph Theoretical Model of Data Structures. PhD thesis, SUNY, Stonybrook, New York.

Vance, D. and Heller, J.

1971 Structure and Content of a Museum Data Bank. CHUM, Vol. 6, no. 2.